

Crime Prediction Using Machine Learning Algorithm

Giridharan S¹, Mr. K. Nirmal²

²Asst. Prof.

^{1,2} KRISHNASAMY COLLEGE OF ENGINEERING AND TECHNOLOGY

Abstract- Crime prediction is of great significance to the formulation of policing strategies and the implementation of crime prevention and control. Machine learning is the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of public property crime from 2015 to 2018 from a section of a large coastal city in the southeast of China as research data to assess the predictive power between several machine learning algorithms. Results based on the historical crime data alone suggest that the LSTM model outperformed KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks. In addition, the built environment data of points of interests (POIs) and urban road network density are input into LSTM model as covariates. It is found that the model with built environment covariates has better prediction effect compared with the original model that is based on historical crime data alone. Therefore, future crime prediction should take advantage of both historical crime data and covariates associated with criminological theories. Not all machine learning algorithms are equally effective in crime prediction.

I. INTRODUCTION

Crime is one of the biggest and dominating problem in society and its prevention is an important task. Daily there are huge numbers of crimes committed. It is required to keep track of all the crimes and maintain a database which may be used for future reference. The current problem we face are maintaining of proper dataset of crime and analysing this data to help in predicting and solving crimes in future. The task is to predict which category of crime is most likely to occur at what place and what time. The objective of this project is to analyse dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, using the technique of machine learning and data science for crime prediction. In our project, we analyse crime data from the city of Vancouver. It consists of crime information like location description, type of crime, date, time, latitude, longitude. The Random Forest algorithm and various other algorithms will be tested for crime prediction and one with better accuracy will be used for training. The objective of this project is to give an idea of how machine learning and analysis of crime can be

used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate.

II. LITERATURE REVIEW

A. The present research examines a structural model of violent crime in Portland, Oregon, exploring spatial patterns of both crime and its covariates. Using standard structural measures drawn from an opportunity framework, the study provides results from a global ordinary least squares model, assumed to fit for all locations within the study area. Geographically weighted regression (GWR) is then introduced as an alternative to such traditional approaches to modeling crime. The GWR procedure estimates a local model, producing a set of mapable parameter estimates and t-values of significance that vary over space.

B. Social networks produce enormous quantity of data. Twitter, a microblogging network, consists of over 230 million active users posting over 500 million tweets every day. Propose to analyze public data from Twitter to predict crime rates. Crime rates have increased in the past recent years. Although crime stoppers are utilizing various technics to reduce crime rates, none of the previous approaches targeted utilizing the language usage (offensive vs. non-offensive) in Tweets as a source of information to predict crime rates. In this paper, hypothesize that analyzing the language usage in tweets is a valid measure to predict crime rates in cities.

C. In recent years the data mining is data analyzing techniques that used to analyze crime data previously stored from various sources to find patterns and trends in crimes. In additional, it can be applied to increase efficiency in solving the crimes faster and also can be applied to automatically notify the crimes. However, there are many data mining techniques.

D. The research presented here has two key objectives. The first is to apply risk terrain modeling (RTM) to forecast the crime of shootings. The risk terrain maps that were produced from RTM use a range of contextual information relevant to the opportunity structure of shootings to estimate risks of future shootings as they are distributed throughout a geography.

E. Introduce a family of models to describe the spatio-temporal dynamics of criminal activity. It is argued here that with a minimal set of mechanisms corresponding to elements that are basic in the study of crime, one can observe the formation of hot spots. By analyzing the simplest versions of our model, we exhibit a self-organized critical state of illegal activities that we propose to call a warm spot or a tepid milieu² depending on the context.

III. PROPOSED SYSTEM

Crime hotspot prediction aims to predict the likely location of future crime events and hotspots where the future events would concentrate. In this paper, random forest algorithm is used for crime prediction.

The randomness of random forest is reflected in two aspects: one is to randomly select the training sample set by using bagging algorithm; the other is to randomly select the split attribute set. Assuming that the training sample has M attributes in total, we specify an attribute number $F < M$, in each internal node, randomly select F attributes from M attributes as the split attribute set, and take the best split mode of the f attributes Split the nodes. The multi decision tree is made up of random forest, and the final classification result is determined by the vote of tree classifier.

The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The Random Forest will be used for crime prediction.

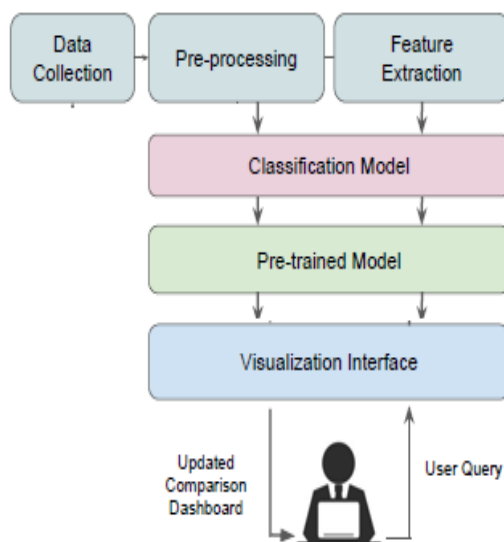


Figure. 1 System Architecture

IV. MODULES

Data Collection: This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the better data and the better model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. Comparison of Machine Learning Algorithms for Predicting Crime Hotspots taken from kaggle and some other source

Dataset: The dataset consists of 821 individual data. There are 27 columns in the dataset, which are described below.

State: State in India

District: District in the state of India.

Year: 2001-2018

Murder: Total number of murder rate

Rape: Total number of rape rate

Theft: Total number of theft rate

Total crime: Total number of total crime rate

Data Preparation: It will transform the data. By getting rid of missing data and removing some columns. First, to create a list of column names that want to keep or retain. Next drop or remove all columns except for the columns that we want to retain. Finally drop or remove the rows that have missing values from the data set.

Model Selection: While creating a machine learning model, need two data set, one for training and other for testing. But now it have only one. So let's split this in two with a ratio of 80:20. It will also divide the dataframe into feature column and label column. Here imported `train_test_split` function of sklearn. Then use it to split the dataset. Also, `test_size = 0.2`, it makes the split with 80% as train dataset and 20% as test dataset. The `random_state` parameter seeds random number generator that helps to split the dataset.

The function returns four datasets. Labelled them as `train_x`, `train_y`, `test_x`, `test_y`. If we see shape of this datasets we can see the split of dataset. It will use Random Forest Classifier, which fits multiple decision tree to the data. Finally train the model by passing `train_x`, `train_y` to the `fit` method. Once the model is trained, it need to Test the model. For that we will pass `test_x` to the predict method. Random Forest is one of the most powerful methods that is used in machine learning for regression problems. The random forest comes in the category of the supervised regressor algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the regressor.

Analyze and Prediction:

In the actual dataset, we chose only 3 features :

State: State in India

District: District in the state of India.

Year: 2001-2018

Prediction :

1. Total number of murder rate
2. Total number of rape rate
3. Total number of theft rate
4. Total number of total crime rate

Saving the Trained Model: Once you're confident enough to take trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle. Make sure you have pickle installed in this environment. Next, let's import the module and dump the model into .pkl file

V. CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with Good Accuracy. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. Generated many graphs and found interesting statistics that helped in understanding Indian crimes datasets that can help in capturing the factors that can help in keeping society safe.

REFERENCES

- [1] M. Felson and E. Poulsen, "Simple indicators of crime by time of day," *Int. J. Forecasting*, vol. 19, no. 4, pp. 595601, Oct. 2003.
- [2] Enhancing Urban Safety and Security. Global Report on Human Settlements 2007, UN-Habitat, Nairobi, Kenya, 2007.
- [3] G. Owusu, C. Wrigley-Asante, M. Oteng-Ababio, and A. Y. Owusu, "Crime prevention through environmental design (CPTED) and built environmental manifestations in Accra and Kumasi, Ghana," *Crime Prevention Community Saf.*, vol. 17, no. 4, pp. 249269, Nov. 2015.
- [4] Y. Wenhao and A. Tinghua, "The visualization and analysis of POI features under network space supported by kernel density estimation," *Acta Geodaetica et Cartographica Sinica*, vol. 44, no. 1, pp. 8290, 2015.
- [5] G. Song, L. Xiao, S. Zhou, D. Long, S. Zhou, and K. Liu, "Impact of residents' routine activities on the spatial-temporal pattern of theft from person," *Acta Geography Sinica*, vol. 72, no. 2, pp. 356367, 2017.
- [6] L. Lin, D. Fang-Ye, X. Lu-Zi, S. Guang-Wen, and J. C. L. Kai, "The density of various road types and larceny rate: An empirical analysis of ZG city," *Hum. Geography*, vol. 32, no. 6, pp. 3239, 2017.
- [7] C. Xu, L. Liu, and S. H. Zhou, "The comparison of predictive accuracy of crime hotspot density maps with the consideration of the near similarity: A case study of robberies at DP Peninsula," *Scientia Geographica Sinica*, vol. 36, no. 1, pp. 5562, 2016.
- [8] G. Rosser, T. Davies, K. J. Bowers, S. D. Johnson, and T. Cheng, "Predictive crime mapping: Arbitrary grids or street networks," *J. Quantum Criminol.*, vol. 33, no. 3, pp. 569594, 2017.
- [9] A. Rummens, W. Hardyns, and L. Pauwels, "The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context," *Appl. Geography*, vol. 86, pp. 255261, Sep. 2017.
- [10] S. Favarin, "This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy," *Eur. J. Criminol.*, vol. 15, no. 6, pp. 702729, Nov. 2018.